DOCUMENT RESUME

ED 430 009 TM 029 722

AUTHOR	Jones, Lyle V.
TITLE	The Assessment of Student Achievement: The Hundred Years War.
PUB DATE	1999-04-00
NOTE	16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999). Cover title varies slightly.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Achievement Tests; Decision Making; *Educational Assessment; Educational History; Elementary Secondary Education; National Surveys; Standardized Tests; *Test Use
IDENTIFIERS	*High Stakes Tests; *National Assessment of Educational Progress; Tyler (Ralph W)

ABSTRACT

The use of achievement tests for high-stakes decision making is discussed, and some comments are made about Ralph Tyler's contributions to educational assessment and what has resulted from them. The assessment war was raging more than 100 years ago, as historical records of testing student achievement demonstrate. Ralph Tyler, in the 1960s, did his best to bring about a truce in the assessment war by inventing the term "assessment" and suggesting that the focus be on the educational attainments of large numbers of people. The National Assessment of Educational Progress (NAEP) advocated by Tyler called for the assessment of a sample of students, each of whom took only a fraction of the exercises, and none of whom received an individual score. Over the past 30 years these objectives have just barely survived. Among other changes, scores are imputed for each child, and these are averaged for any specific subgroup. Many changes in the NAEP have been supported by psychometric considerations, but some have compromised Tyler's vision by supporting pressures for reported scores for school districts, schools, classrooms, and individual students. Many of the problems evident as testing stakes become higher are identified. It is shortsighted to accept test scores as the ultimate criterion of the benefits of education, and more appropriate criteria must be developed. Some attachments illustrate historical controversy over assessment. (Contains four references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 009

Student Assessment of Student Achievement: The Hundred Years War

Lyle V. Jones

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Lyle Jones
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM029722

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in Montreal Canada
April 1999

BEST COPY AVAILABLE

The Assessment of Student Achievement: The Hundred Years War

Lyle V. Jones

The University of North Carolina at Chapel Hill

I intend to begin with some remarks about the use of achievement tests for high-stakes decision making, and then to comment on Ralph Tyler's contributions to educational assessment and on what has become of them. After referring again to high-stakes testing, I'll suggest an agenda item for future research.

Consider first this comment about high-stakes testing:

[First overhead]

Note the date, 1887. The reference is to the system of "payment by results" for the publicly supported elementary schools in England and Wales. Grants to each school were based on annual inspections that entailed the testing of individual students on reading passages and on arithmetic test cards.

* Invited address, Division D of AERA. Montreal, Canada, April 21, 1999.

Was this a high-stakes testing system? It certainly was for teachers and for school managers, whose salaries were linked to the amount of the grant (Sutherland, 1973). And an Inspectors' visit was an anxiety-provoking event. One report reads as follows:

[Second Overhead]

The inspectors, sub-inspectors, and their assistants weren't having much fun, either.

[Third overhead]

Clearly, the assessment war was waging even earlier than 100 years ago! The political demand for school accountability led many teachers to concentrate on drilling their pupils to prepare them for the tests, at the expense of broader objectives of instruction. (Does that sound familiar now, over a century later?)

Ralph Tyler did his best to effect a truce in an ongoing war. He invented the term assessment to distinguish it from three other forms of educational appraisal: first, testing achievements of individual students to assign grades or to select students for

further opportunities, second, diagnosing learning difficulties of a student (or of a class) to plan subsequent teaching, and third, evaluating the effectiveness of a curriculum or a set of teaching methods. In contrast, Tyler proposed that the focus of assessment be not on individual students, classrooms, schools or school systems. Assessments furnish information about the educational attainments of large numbers of people, perhaps of different ages, different demographic groupings, and different geographic regions. The purpose of assessment is analogous to that of the Gross National Product, or of the Consumer Price Index, or of health and mortality indices, to provide dependable information about population and sub-population change over time, in this case about the progress of education.

As the National Assessment of Educational Progress (NAEP) was designed, only a sample of students would be assessed, no student would take more than a fraction of the exercises, and no score would be obtained from any student's performance. Exercises -- many of them in the form of hands-on problems to be solved -- would represent a broad range of difficulty and a

full array of educational objectives in ten different subject areas. Professional administrators were trained to provide highly controlled assessment conditions. Exercises were read aloud, so that deficiencies in reading would not prevent good performance in math, say, or in citizenship. An "I don't know" alternative was offered to discourage guessing and to reduce non-response. Results of periodic assessments would be reported, exercise by exercise, so that for four different age groups, 3, 9, 17, and young adult, the public would have concrete evidence about what respondents know and can do.

Over the past 30 years, some of Tyler's objectives have survived, but just barely. First, the desired rich variety of exercises was compromised, in favor of more traditional multiple-choice and short-answer items, the kinds of items with which testing companies were familiar. Exercises became quite homogeneous in difficulty, with fewer very easy or very difficult ones. The young-adult sample was eliminated, and school grade has replaced age as the primary unit of assessment. The ten subject areas have received uneven attention, with math,

reading, science, and writing assessed far more often than literature, social studies, art, music, citizenship, and career development. No longer are exercises read aloud, nor has an "I don't know" alternative been retained. For state assessments, local school personnel now administer the exercises, which raises questions about the uniformity of administration.

Instead of reporting a percent-correct score for each exercise, scale scores were developed, for large clusters of exercises. More recently, reporting has been by "achievement levels," so as to compare actual performance with how good performance "should be".

Using IRT technology, scores now are imputed for each child in the sample, even though different children take different sets of exercises. Imputed scores then are averaged for any specified subgroup of children.

Many of these changes were well-intentioned, and some clearly are supported by psychometric considerations and by the need to better communicate results to the public. Nonetheless, some changes have compromised Tyler's vision of

assessment. For, along with the changes have come pressures to report scores not just for large subpopulations, but by school district, by school, by classroom, and for each individual child. Indeed, President Clinton and the U.S. Department of Education promote "a personalized version of NAEP" (Riley, 1997) for every child, a voluntary national test derived from the NAEP model. And most states have adopted or are developing high-stakes tests in reading and math, often to be used as a basis for promotion from grade-to-grade, for receipt of a high school diploma, and for salary supplements to teachers -- shades of 1887!

Numerous unresolved problems attend these procedures. What accommodations should be provided for students who are visually handicapped or in special education classes? What special efforts will be dedicated to helping students who fail to pass the tests? Will the testing programs lead to increased levels of school dropout thereby helping some students at the expense of others? Will teachers focus on preparing students

for the tests at the expense of failing to meet other important educational objectives?

Illustrative of some of these problems is a school principal's recent announcement to teachers at a middle school near my home in North Carolina. The school goal is to elevate end-of-grade test scores in reading and math so that by 2003, 95% of the students will score at or above grade level. Between now and then, students scoring below grade level will not be permitted to enroll in elective courses, but will be assigned to remedial math and reading classes instead.

As one might expect, the teachers of drama, art, dance, and music are appalled. They cite many instances of students whose engagement in their subjects result in a heightened commitment to school and increased performance in core subjects as well.

[Fourth Overhead]

Quite obviously, good education can be both, instilling knowledge, but also lighting fires, encouraging engagement. And neither goal is sufficient without the other. For accountability, an imbalance has been created because we have

learned to measure achievement much better than we can measure engagement. Isn't it time to develop measures of the latter?

A recent study by Mahoney & Cairns (1997) supports the importance of engagement for reducing high school dropout, especially for at-risk students. In a longitudinal study, the authors followed about 400 students through middle school and high school. Students were classified as highly competent, marginally competent, or at risk, based upon annual ratings by teachers of each student's academic and personal competence. Engagement was measured by the extent of extracurricular involvement in athletics, student government, student clubs, music or drama, journalism, and school assistantships. For all three competence groups, but most dramatically for at-risk students, the proportion of dropouts in high school declines for students with higher levels of extracurricular involvement.

Such promising results suggest that devising systematic measures of student engagement may have a considerable payoff. Adding measures of school engagement to measures of

school achievement could help to restore a balance between two important components of effective schooling.

Perhaps the most serious impediment to educational reform is confusion about the criterion. For many states -- and perhaps for the nation as well -- test scores seem to be accepted as the criterion for effective education. So the message to teachers is to train students to elevate their test scores. How shortsighted it is to accept test scores as an ultimate criterion of the benefits of education. At best, tests can serve to predict one component of successful schooling. By establishing a richer and more appropriate criterion, we would be challenged to develop richer and more appropriate predictors, certainly to include indices of commitment to learning -- that is, the lighting of fires -- as well as indicators of what students have learned.

References

Mahoney, J.L. & Cairns, R.B. (1997). Do extracurricular activities protect against early school dropout? Developmental Psychology, 33, 241-253.

Riley, R. W. (1997). Testimony of Secretary Richard W. Riley. Statement to the Senate Labor, Health & Human Services & Education Subcommittee of the Senate Appropriations Committee, September 4.

Sutherland, G. (1973). Policy-making in elementary education 1870-1895. London: Oxford University Press.

Tyler, R.W. (1966). The objectives and plans for a National Assessment of Educational Progress. Journal of Educational Measurement, 3, no. 1, 1-4.

An elementary school principal comments on high-stakes student testing:

"... a teacher knows that his whole professional status depends on the results he produces and he really is turned into a machine for producing those results; that is, I think, unaccompanied by any substantial gain to the whole cause of education."

*Archdeacon Sir Lovelace Stamer (1887) Minutes of the Cross Commission. London, Parliamentary Papers. Cited in Sutherland (1973).

"Two inspectors came once a year and carried out a dramatic examination. The schoolmaster came into school in his best suit; all the pupils and teachers would be listening till at ten o'clock a dog-cart would be heard on the road even though it was eighty yards away. In would come two gentlemen with a deportment of high authority with rich voices. Each would sit at a desk and the children would be called in turn to one or other. The master hovered round, calling children out as they were needed. The children could see him start with vexation as a good pupil stuck at a word in the reading book he had been using all year, or sat motionless with his sum in front of him. The master's anxiety was deep, for his earnings depended on the children's work. One year the atmosphere of anxiety so affected the children that, one after another as they were brought to the Inspector, the boys howled and the girls whimpered. It took hours to get through them."

"I feel as if I were writing a despatch in the midst of a battle, so dire is the din of educational conflict around."

Her Majesty's Inspector Temple, 1874

"Education is not the filling of a pail, but the lighting of a fire."

--William Butler Yeats